

# Hierarchical structure and time-lag correlation in Worldwide Financial Markets

Zeyu Zheng<sup>1</sup> and Kazuko Yamasaki<sup>1,2</sup>

<sup>1</sup>*Department of Environmental Sciences,  
Tokyo University of Information Sciences, Chiba 265-8501, Japan*

<sup>2</sup>*Center for Polymer Studies and Department of Physics,  
Boston University, Boston, MA 02215, USA*

## Abstract

Recently, many studies indicated that the minimum spanning tree (MST) network whose metric distance is defined by using correlation coefficients have strong implications on extracting information from return time series. However in many cases researchers may hope to investigate the strength of interactions but not the directions of them. In order to study the strength of interaction and connection of financial asset returns we propose a modified minimum spanning tree network whose metric distance is defined from absolute cross-correlation coefficients. We had investigated 69 daily financial time series, which constituted by 3 types finance assets (29 stock market indicator time series, 21 currency futures price time series and 19 commodity futures price time series). Empirical analyses show that the MST network of returns is time-dependent in overall structure, while same type financial assets usually keep stable inter-connections. Moreover each asset in same group show similar economic characters. In other words, each group concerned with one kind of traditional financial commodity. In addition, we find the time-lag between stock market indicator volatility time series and EUA (EU allowances), WTI (West Texas Intermediate) volatility time series. The peak of cross-correlation function of volatility time series between EUA (or WTI) and stock market indicators show a significant time shift ( $> 20days$ ) from 0.

PACS numbers: PACS numbers:89.65.Gh, 89.20.-a, 02.50.Ey

In study of complex system, the selection of statistically reliable information from correlation structure is a very useful method [1–12]. This process usually be termed with the locution "filtering procedure" as in [8]. Useful examples of filtering procedure based on correlation matrix of return time series are the hierarchical clustering [1–7], procedures based on the random matrix theory [13–18], networks from minimum spanning tree [1–5]. Correlation structure study not only limited in stock return time series [1–3], quasi-synchronously recorded time series of worldwide stock exchanges market index [2], and volatility increments of stock return time series [7] are also be studied.

Financial record time series include not only stock price time series but also many other types, such as features price, treasury yield, market index and so on. We believed that investigate the relationship of multi-type quasi-synchronously financial time records can indicate the physical interdependent relationship of market or commodities that the financial asset reflected. Moreover a relationship map of financial assets can help us figure out the movement of speculative capital. In other words, we hope our study can reveal and separate the effect from speculative capital and the inherent characteristics of the asset itself.

We had investigated 69 daily financial time series during the years from January 2007 to September 2011. The data set include 21 currency futures price time series, 19 commodity futures price time series which are taken from <http://data.theice.com/ViewData/Default.aspx> and 29 stock market indicator time series, which are taken from <http://finance.yahoo.com>.

Additionally, we should point out that trading may occur at different time in two different cities implies that some markets are open during the time whereas others are closed, for example the New York and Tokyo stock markets. The analysis of daily data of closure values may induce spurious correlations introduced just by the specific time at which the records are stored. The effect of non-synchronous trading in time series analysis had been well stated Ref [20–22]. In fact, the highest degree of correlation between different markets may be detected in one day time-lag because of time difference. For example, the highest correlation is observed between the closure return series of the New-York stock exchange at day  $t$  and the closure return series of the Tokyo stock market at day  $t + 1$  [2, 3]. Since the time difference will no more than one day in the earth, the time-lag would not larger than one day too.

Recently, a few papers have revealed that the correlation structure which is described by the ultrametric space and the hierarchical organization is informative for financial return

time series. This is obtained by defined a metric distance that is defined as

$$d_{ij} = \sqrt{2(1 - \rho_{ij})} \quad (1)$$

in each pair of elements  $i$  and  $j$ . With this distance  $d_{ij}$  fulfills the three axioms of metric: i)  $d_{ij} = 0$  if and only if  $i = j$ ; ii)  $d_{ij} = d_{ji}$  and iii)  $d_{ij} \leq d_{ik} + d_{kj}$  [1-3].

In this study, we focus on investigate the strength of connection that is described as distance between each pair return time series, no matter correlation or anti-correlation big correlation magnitude indicate strong connection, on the contrast small correlation magnitude indicate weak connection. For this reason, we use absolute Pearson correlation coefficient instead of Pearson correlation coefficient in equation 1. A new correlation coefficient based distance equation is defined as:

$$d_{ij} = \sqrt{2(1 - |\rho_{ij}|)} \quad (2)$$

with the  $\rho_{ij}$  is defined as correlation coefficient of assets  $i$  and  $j$ . This equation also fulfills three axioms of a metric distance. The first and second axioms are easily verified because  $\rho_{ij} = 1$  implies  $d_{ij} = 0$ , while  $\rho_{ij} = \rho_{ji}$  implies  $d_{ij} = d_{ji}$ . For the validity of axiom (iii), consider  $n$  scaled time series  $Y_1, Y_2, \dots, Y_n$  and a single scaled time series  $X$ , which have means of 0 and standard deviation of 1. All time series have same length. Since  $-\rho_{X, Y_i} = \overline{(-Y_i) \times X}$  where  $i \in n$ , it is always possible that we can create  $n$  new time series which satisfy  $\rho'_{X, Y'_i} \geq 0$  as  $i \in n$ . Then according to the definition of correlation coefficient,  $\left. \begin{matrix} \rho'_{X, Y'_i} \geq 0 \\ \rho'_{X, Y'_j} \geq 0 \end{matrix} \right\} \Rightarrow \rho'_{Y'_i, Y'_j} \geq 0$  we can conclude that  $\rho'_{Y'_i, Y'_j} = |\rho_{Y_i, Y_j}| \geq 0$  are satisfied for any  $i, j$  when  $i, j \in n$ . So  $d_{ij} = \sqrt{2(1 - |\rho_{ij}|)}$  can be rewrite as  $d_{ij} = \sqrt{2(1 - \rho'_{ij})}$ , the equation is the same as equation 1 that satisfy the third axiom. Therefore  $d_{ij} = \sqrt{2(1 - |\rho_{ij}|)}$  are also fulfills the three axioms of a metric distance.

For each of the 22 financial time series, we calculate the return time series that is the change of logarithmic price of time series  $i$  as equation (5).

$$R_i(t) \equiv \ln(Y_i(t+1)) - \ln(Y_i(t)) \quad (3)$$

Here  $Y_i(t)$  is the daily price time series of financial asset  $i$  in day  $t$ . For each of the 22 time series, we also calculate the volatility time series which is defined as absolute values  $|R_i|$ .

$$V_i(t) = |R_i|(t) \equiv |\ln(Y_i(t+1)) - \ln(Y_i(t))| \quad (4)$$

Consider two time series  $\{y_t\}$  and  $\{y'_t\}$ , where  $t = 1, 2, \dots, N$ . We can define the cross-correlation function between  $\{y_t\}$  and  $\{y'_t\}$  as:

$$C_{y,y'}(n) \equiv \overline{(y_t - \mu)(y'_{t+n} - \mu')} / (\sigma\sigma') \quad (5)$$

where  $\mu$  is the mean and  $\sigma$  is the standard deviation of series  $\{y_i\}$ , while  $\mu'$  is the mean and  $\sigma'$  is the standard deviation of series  $\{y'_i\}$ .

As a basic theory of econophysics, it is believed that the long-rang memory can not exist in any return time series from an efficient market. Suppose that the long-range auto-correlation is exist in a return time series, the investors may obtain benefits by using the information of long-rang memory, which is contradiction to the fact of efficient market [9]. Consider the cross-correlation function (5) between return time series of asset  $i$  and asset  $j$ , any significant cross-correlations  $C_{R_i,R_j}(n)$  in  $n \neq 0$  of two return time series also means the hypothesis of efficient market is wrong. Therefore the significant cross-correlation  $C_{R_i,R_j}(n)$  will only exist as  $n = 0$ . However because of the time-lag between different market in different cities, significant cross-correlation  $C_{R_i,R_j}(n)$  may also exist as  $n = -1$  or  $n = 1$ . Additionally we only care about he magnitude not the sign of cross-correlation. So, it is reasonable that we define absolute correlation coefficient as

$$\rho_{ij} = \text{Max}(|C_{R_i,R_j}(n)|) \quad (6)$$

while  $n = -1, 0, 1$ .  $\rho_{ij}$  can describe the magnitude of the two quasi-synchronous return time series of asset  $i$  and asset  $j$ , which like the Pearson correlation coefficient of two synchronous return time series.

On the other hand, for volatility time series, because the existence of long-range cross correlation  $C_{V_i,V_j}(n)$  [9–12], moreover the time-lag of significant cross-correlation can not help investor to obtain benefits. Therefore the significant cross-correlation  $C_{V_i,V_j}(n)$  may exist while  $n \gg 0$  or  $n \ll 0$ . As a sample, the correlation functions of volatility  $C_{V_i,V_j}(n)$  and return  $C_{R_i,R_j}(n)$  between FTSE100 and NYdow are shown in Fig1. The peaks (highest correlations) of two correlation functions are all near 0. But the correlation function of returns shows a fast-decaying,  $C_{R_i,R_j}(n) \approx 0$  when  $n \neq 0$ , On the contrast the correlation function of volatilities shows a slow-decaying,  $C_{V_i,V_j}(n) > 0$  when  $n > -50$  and  $n < 50$ .

In the rest of the letter, we will firstly show the stability and structure of MSTs which are made by using the distance based on absolute correlation coefficient, and it can indicate

the interaction correlation of financial return time series, then we will show the correlation function graphs of volatility that indicate the relationship of cross-correlation coefficient  $C(n)$  and time-lag  $n$ , especially we focus on the value of time-lag  $n$  as LOWESS (locally weighted scatter plot smoothing) values of  $C(n)$  equal to its maximum value (highest correlations).

In fig2, we find that however the MSTs show significant different structures in different calendar years, same type of finance asset usually gather together. Even in the year of 2007 development of the subprime crisis and 2008 global financial crisis, such clustering and connection inside in same type financial assets time series are not be disturbed. It indicate some very strong and stability connection which mostly come from the cross-correlation based on basic economic features and interactions, should be exist in these return time series which reflect same type economic group. These connections are stability to time, and little affect by business conditions. Moreover we also find that the blue (stock market indicators) and green (currency futures) groups show stronger inter-connections than red (Commodities) group. For stock market indicators and currency futures, in most of the time the financial factors may be the only important reason of price changes. On the other hand, for commodity futures, there may have other reason of price changes that is the contradiction between supply and demand. The contradictions between supply and demand are changed following the type of commodity and calender years, that decrease the stability of MST. According to the hypothesis, if we increase the time span of time series for cross-correlation, the connection should become more stable. In Fig3, we show the MST by using the time series from January 2007 to September 2011, only two Coal futures are not connect with commodity group. It shows more strong inter-connection than single year time series.

Furthermore we find that EUA (European Union allowances) futures mostly (except in 2008) connect with the base electricity and nature gas futures, which show stability correlations among them. Since Power generation accounts for about one-quarter of total emissions of carbon dioxide, and nature gas is the most resource of electricity generation in UK, the stability connections of EUA, UK base electricity future, and UK nature gas future in our MST graph is reasonable. It reflect such economic relationships.

In fig 4, we describe the cross-correlation functions of volatility time series. We show the cross-correlation functions of mainly stock market with EUA in (a), with WTI in (b). It is easy to be found that a systemic time shift is existed between EUA, WTI and stock market indicator time series. The maximum of cross-correlation coefficients in most function are

close or bigger than 0.2, it indicate not so strong but definite correlations of stock market and EUA or WTI.

The correlation function of volatility is a slow-decaying function. It is much slower than the correlation function of return time series see Fig1, that means a long-rang cross-correlation relationship is existed. If we assume significant cross-correlation between different volatility time series as an information transformation of different assets. The shift of highest values can be considered as a time-lag between information transformations. It is worth pointing out that the time-lag between each pair of stock markets is approximately 0. We showed such time-lag simply in fig 4 (c). For example, there have some information are send out from stock markets, or some information affect stock markets in time of 0 day. Then EUA and WTI futures will be affected by these information roughly after 30 and 90 days respectively.

In this paper, we analyzed the correlation function of return and volatility time series, construct the MST based on return time series, and find out the time-lag in correlation function of volatility. From these analysis we get some results: i) The stability of MST structure inside groups each of them concerned with one kind of traditional financial commodity. This phenomenon reflects the stability of the basic rule of economic activity, the interaction between economic time series is not easily affected by the capital movement. The method of absolute cross-correlation coefficient based MST has strong implications on reveal the ongoing debate about the relationships of different financial commodity time series. ii) We find the time-lag of correlation function of volatility time series exist in the stock markets and EUA, WTI markets. The time-lag indicate that there may have systemic difference of spread volatility of economic information while it active on different financial assets. This results provide us an new approach that can predict the financial risks in much longer time interval.

- 
- [1] R. N. Mantegna, Eur. Phys. J. B **11**, 193-197 (1999).
  - [2] G. Bonanno, N. Vandewalle, and R.N. Mantegna, Phys. Rev. E **62**, R7615-R7618 (2000).
  - [3] G. Bonanno, F. Lillo, and R.N. Mantegna, Quantitative Finance **1**, 96-104 (2001).
  - [4] J.-P. Onnela, K. Kaski and J. Kertesz, Eur. Phys. J. B **38**, 353-362 (2004).

- [5] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, and A. Kanto, Phys. Rev. E **68**, 056110 (2003)
- [6] D. Song, M. Tumminello, W. Zhou, and R. N. Mantegna, Phys. Rev. E **84**, 026108 (2011)
- [7] S. Miccichea, G. Bonannob, F. Lilloa, and R. N. Mantegna, Physica A **324**, 66-73, (2003)
- [8] M. Tumminello, F. Lillo, and R. N. Mantegna, J. Econ. Behav. Org. **75** 40 (2010)
- [9] R. Mantegna and H. E. Stanley, An Introduction to Econophysics Correlations and Complexity in Finance, Cambridge University Press (2000).
- [10] K. Yamasaki, L. Muchnik, S. Havlin, A. Bunde, and H. E. Stanley, Proc. Natl. Acad. Sci. USA **102**, 9424 (2005).
- [11] F. Wang, K. Yamasaki, S. Havlin, and H. E. Stanley, Phys. Rev. **73**, 26117 (2006).
- [12] F. Wang, K. Yamasaki, S. Havlin, and H. E. Stanley, Phys. Rev. E **77**, 016109 (2008).
- [13] L. Laloux, P. Cizeau, J.-P. Bouchaud and M. Potters Phys. Rev. Lett. **83** 1467 (1999).
- [14] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, Phys. Rev. Lett. **83** 1471 (1999).
- [15] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, Phys. Rev. E **65** 066126 (2002).
- [16] A. Utsugi, K. Ino, and M. Oshikawa Phys. Rev. E **70** 026110 (2004).
- [17] B. Podobnik, D. Wang, D. Horvatic, I. Grosse, and H. E. Stanley, EPL **90** 68001 (2010).
- [18] D. Wang, B. Podobnik, D. Horvatic, I. Grosse, and H. E. Stanley, Phys. Rev. E **83** 046121 (2011).
- [19] S. R. Stavins, Handbook of Environmental Economics, Elsevier, edition 1, volume 1 (2001)
- [20] A. W. Lo, and A. C. Mackinlay, Journal of Econometrics **45** 181 (1990)
- [21] K.G. Becker, J. E. Finnerty, and M. Gupta, Journal of Finance **XLV** 1297 (1990)
- [22] W.-L. Lin, R. F. Engle, and T. Ito, The Review of Financial Studies **7** 507 (1994)

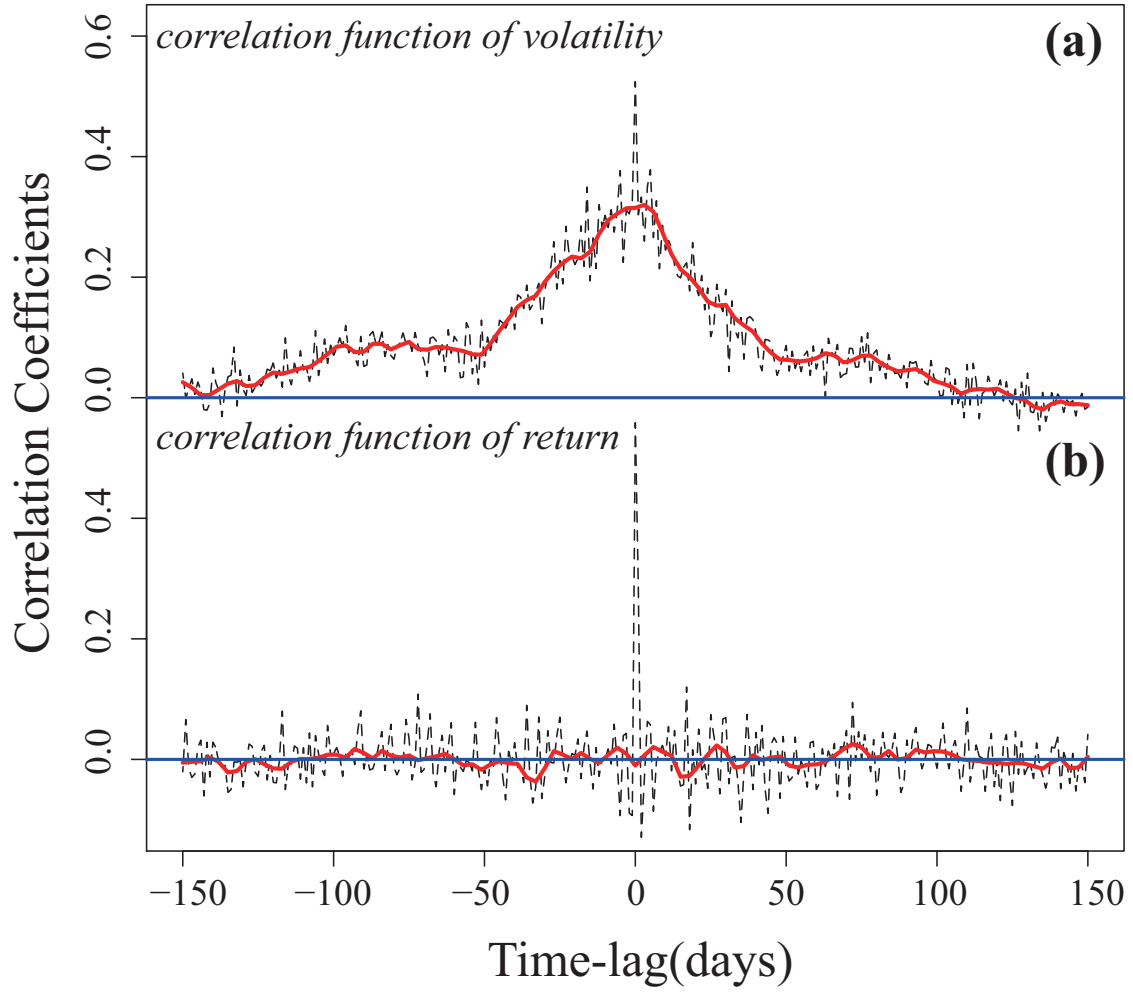


FIG. 1: The cross-correlation function  $C(n)$  of volatility (a) and return (b) time series between NYdow and FTSE100. Both one show the maximum and significant correlation coefficient near time-lag 0 (dotted curve). Solid lines show the LOWESS (locally weighted scatter plot smoothing) values of  $C(n)$ , its smoother span is 30 days.



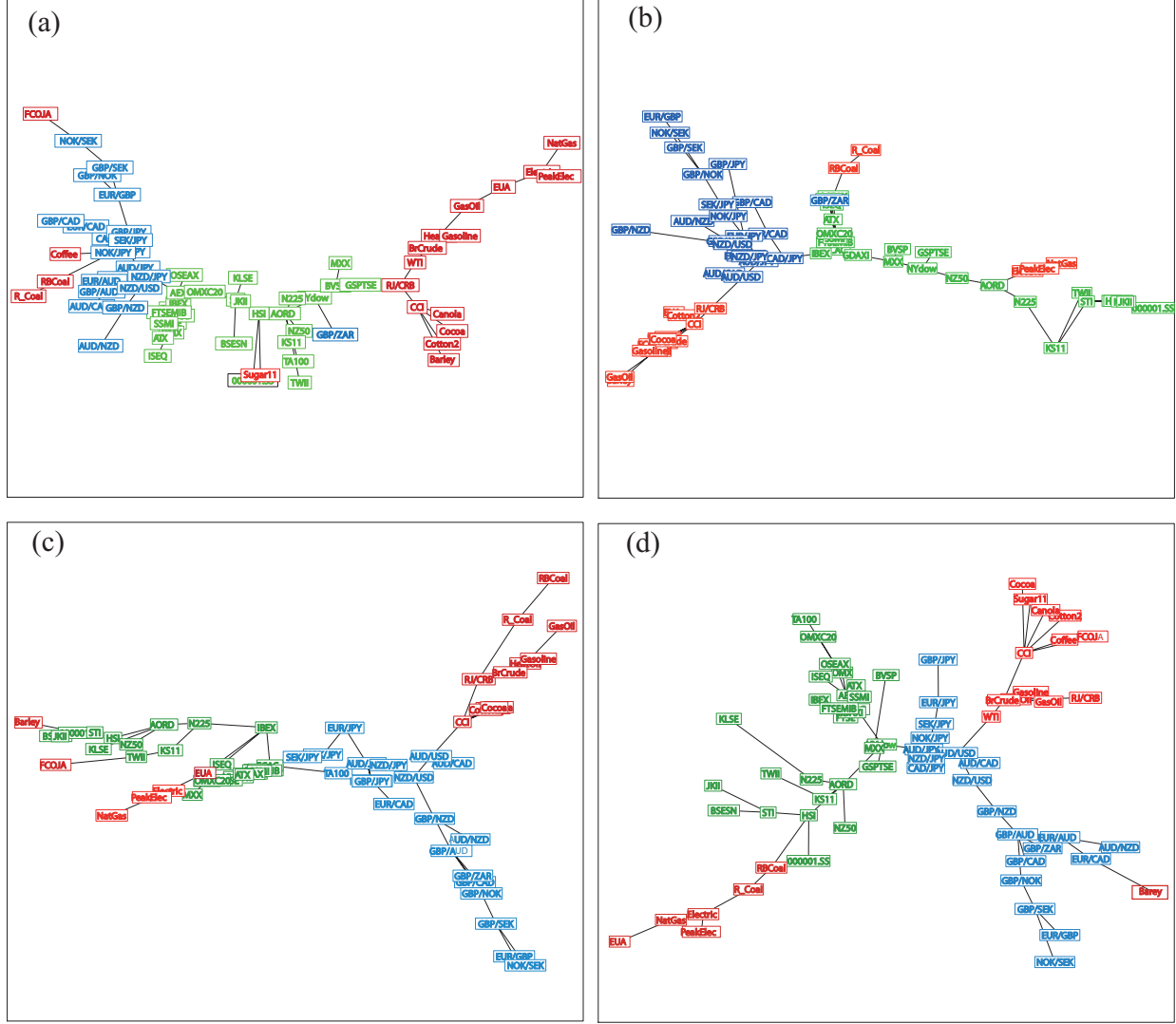


FIG. 2: The minimal-spanning tree (MST) obtained from the absolute correlation coefficients  $|\rho_{ij}|$  of the set of 69 return daily data during in individual calendar years (a) 2007 (b) 2008 (c) 2009 (d) 2010. Red indicates Commodity futures, blue and green indicate currency futures and stock market indicators respectively.

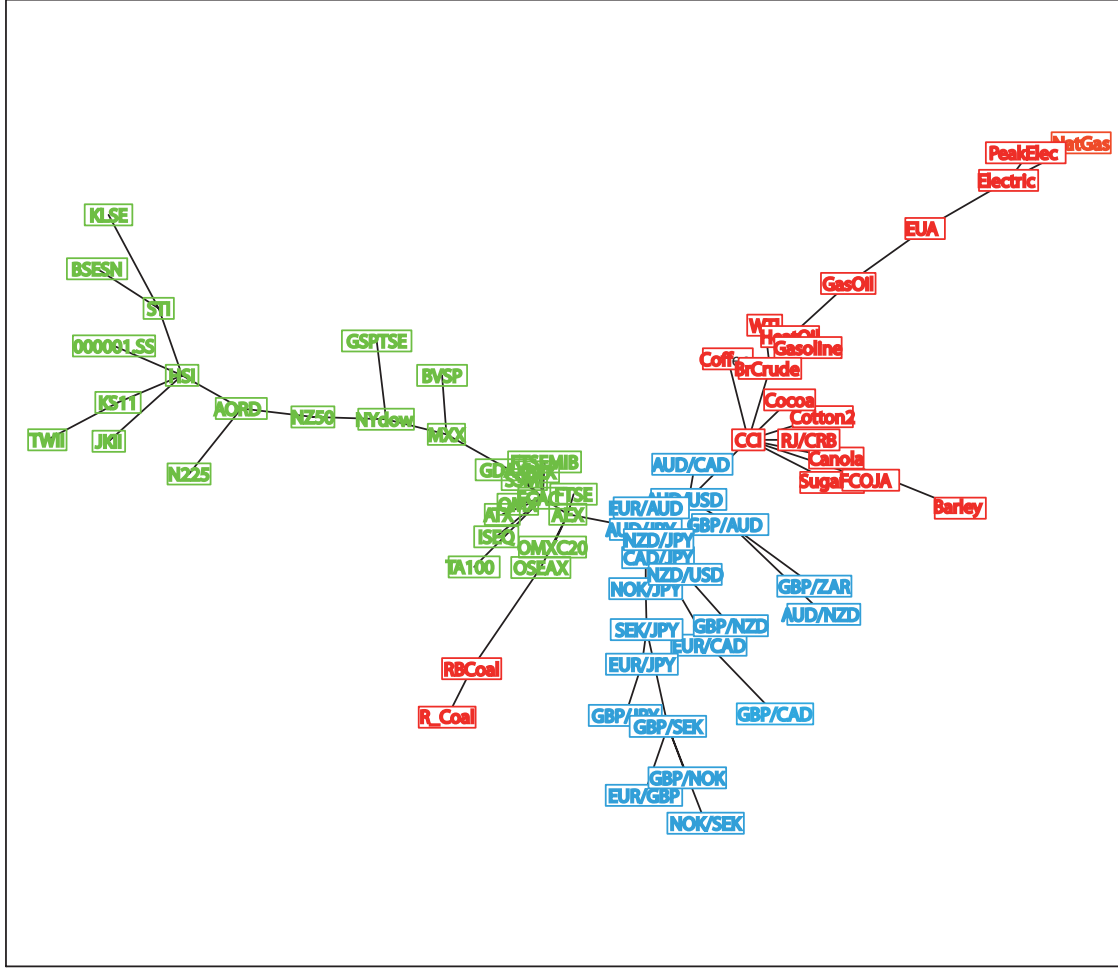


FIG. 3: The minimal-spanning tree (MST) like in fig2 during January 2007 to September 2011.

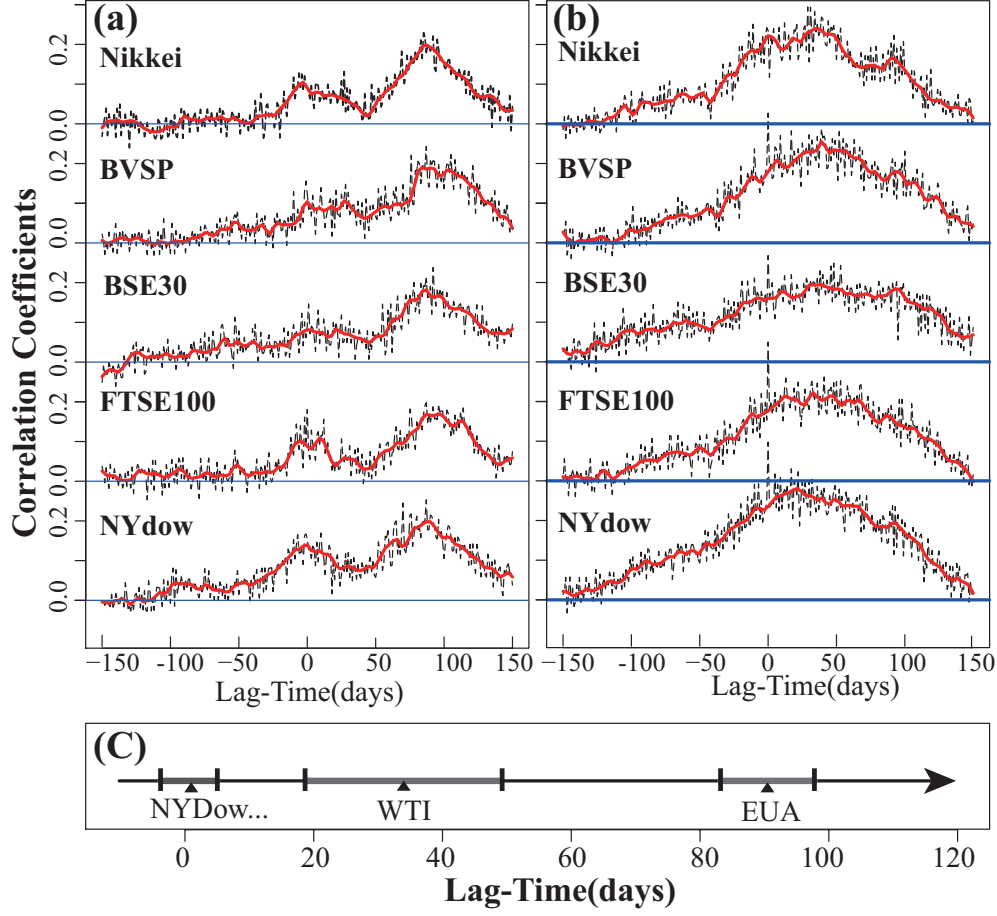


FIG. 4: Cross-correlation Function  $C(n)$  of volatility daily time series between (a) EUA , (b) WTI and 5 main Stock market indexes in the world. Red lines indicate the locally weighted scatter plot smoothing values (LOWESS) of  $C(n)$ . Graphs show systemic time shift of highest cross-correlation function. For most stock market, such systemic time shift of cross-correlation function can be observed. (c) indicate the average time-lag between EUA, WTI and Stock market indicators, error bars show the standard deviations. The lag-time(day) are calculated from the lag-time  $n$  of highest LOWESS that greater than 0.15.